INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 05 ISSUE: 06 | JUNE - 2021 ISSN: 2582-3930

Sports Video Classification

First Author Name, vijay v nikam Second Author Name, Samyak Badjate Third Author Name. shubham patil Fourth Author Name. Shubham Junare Fifth Author Name. Nilesh Patil

***Abstract*— Sports classification has considerable importance for digital content archiving in broadcasting companies. The considered classification categories (topics) are either general (talk show, sport, movie, cartoon ...) or more specific (summer and winter Olympic sports, e.g. cycling, tennis, archery, box ...). At first, each frame of the video is classified separately. In this work, deep neural networks are used, combining convolutional and recurrent networks to classify 15 individual sports classes. It is also a subdivision of human action recognition, which further contributes to understand the context of video scenes. The sports dataset is hand-crafted to focus on sports action based classification. CNN extracted features are combined with temporal information from RNN to formulate the general model to solve the problem..**

***Index Terms*— Sports Classification, Transfer Learning, CNN, RNN.**

I. INTRODUCTION

Growing amount of multimedia data requires sophisticated algorithms for its effective processing, e.g. automatic speech recognition for audio data and computer vision for image and video data. Such algorithms can be used for multimedia data indexation. The goal of this paper is to compare several feature extraction methods and classifiers in the task of image scene classification (or topic classification). The task is applied not only to standalone images, but to continuous videos, without apriori knowledge about the moments when the topic changes. Additionally, we compared online and offline usage of our system. The intended use for the online system is sport classification in real-time, used in automatic speech recognition system that switches language models depending on the current topic that is present in the broadcast.

In different mediums of broadcasting such as television, internet, streaming services etc. sports are a major section and thousands to millions of sports videos flood the data servers daily. It is crucial to index individual sport based on their category if we want to further process them, like analysis of the match, forming new tactics for coaches. Additionally, broadcasting companies can manage their activity easier, while searching specific types of videos without tedious manual work. Since a video sequence is analyzed to classify it, this work is related to the analysis of the context of a scene; thus, it can be further applied to advance high-level understanding of video scenes in machines. There have been many attempts to classify video information with respect to the context of the scenes, in recent times DNN or deep neural network-based models are widely used and quite effective for solving complex tasks of computer vision and signal processing fields. To recognize any sports, human beings only need to consider a set of actions, sometimes the surrounding elements are also taken into consideration. To build a system which can classify multiple sports classes based on sequential image information. Human intuition is the motivation behind the proposed methodology, meaning the system needs to consider the image information simultaneously in the spatial and temporal domain. Convolutional Neural Networks (CNN) are powerful spatial feature extractors while Recurrent Neural Networks (RNN) use internal memory gates to process information through time, by combining these two models we analyze sports video sequences in different combinations and present the experimental results.

II. Literature Survey

Varying methodologies have been developed to classify sports from different types of inputs such as image, audio, video etc. Authors in [1] develop three methods – two kinds of neural nets and texture code cue. Also, their combination was used, and the results were compared, the best performing method was the neural net cue. In [2], the authors used thermal imaging, heatmaps were produced and projected into a low dimensional space using two different techniques - PCA and Fischer's Linear Discriminant. Their overall result was promising for 5 classes. Audio and video features were combined for sports classification in [3], authors extract MFCC features from audio and visual motion features, the results were satisfactory. From continuous tv broadcasts of any videos [4] attempts to classify sports among different kinds of non-sports classes, for feature extraction CNN fc7 + PCA 200 was applied and SVM was used as the classifier. They claim to have achieved good results. Whereas, in [5] the authors try to do sports classification in mobile videos. They consider three kinds of data as inputs – sensor, audio. For classification, they develop multiclass-SVM. They also consider spatial and temporal analysis. A comprehensive experimental analysis was given of their different fusion combinations. In [6], authors used Hidden Markov Models (HMM) to classify specific events inside sports videos. But they do not give details information about accuracy, only the computation time was given.

III. Related Work

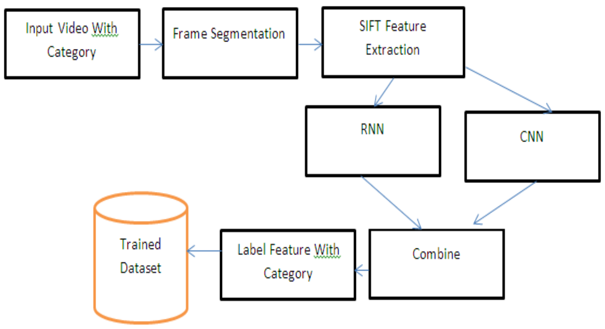
The task of scene classification (or topic classification, genre classification) was often based on low level image analysis using color histograms, motion rates etc. Such low level image features cannot describe high level concepts and semantics that are present in the image. Currently one of the best-performing high-level feature extraction methods is based on deep convolutional neural networks. Implementation of such a network in is already trained on 1.3 million images from 1,000 classes. Because these 1,000 classes can differ from own desired topic classes, the options are either to retrain the neural network with own classes, which requires large amount of data,

 ISSN: 2582-3930

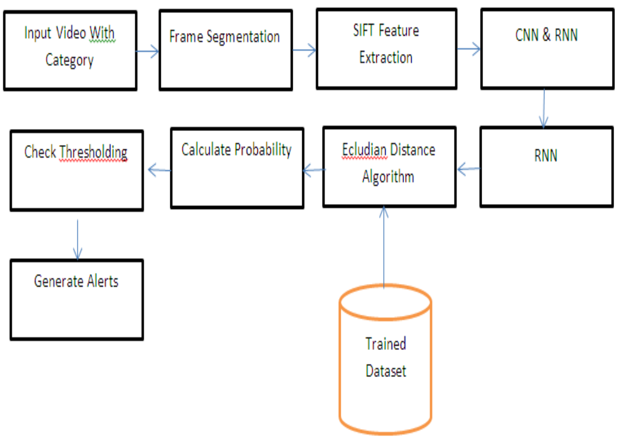
or the neural network can be used only as a feature extractor and own classifier for own classes can be trained based on the features extracted from the existing neural network that was trained on different classes. Such features are proven to be high-level, being related to concepts and semantics rather than to low level features like colors or edges.

IV. Proposed Approach

Training Module:



Testing Module:



It is crucial to index individual sport based on their category if we want to further process them, like analysis of the match, forming new tactics for coaches. Additionally, broadcasting companies can manage their activity easier, while searching specific types of videos without tedious manual work. Since a video sequence is analyzed to classify it, this work is related to the analysis of the context of a scene; thus, it can be further applied to advance high-level understanding of video scenes in machines.

There have been many attempts to classify video information with respect to the context of the scenes, in recent times DNN or deep neural network-based models are widely used and quite effective for solving complex tasks of computer vision and signal processing fields. To recognize any sports, human beings only need to consider a set of

actions, sometimes the surrounding elements are also taken into consideration. To build a system which can classify multiple sports classes based on sequential image information. Human intuition is the motivation behind the proposed methodology, meaning the system needs to consider the image information simultaneously in the spatial and temporal domain. Convolutional Neural Networks (CNN) are powerful spatial feature extractors while Recurrent Neural Networks (RNN) use internal memory gates to process information through time, by combining these two models we analyze sports video sequences in different combinations and present the experimental results.

V. CONCLUSION

We have proposed and evaluated system for topic classification from video data, based on single image classification. The system can be used both in real-time and offline applications, performing slightly better in the second case. We proposed an AlexNet convolutional neural network-based model for shot classification of field sports videos. Our framework is robust to camera variations, scene change, action speeds.

REFERENCES

[1] K. Messer, W. Christmas, J. Kittler, "Automatic sports classification", Proc. IEEE Int. Conf. Pattern Recognition, pp. 1005-1008, 2002.
[2] R. Gade, T. Moeslund, "Sports type classification using signature heatmaps", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 999-1004, 2013.
[3] R. Gade, M. Abou-Zleikha, M. G. Christensen, T. B. Moeslund, "Audio-visual classification of sports types", 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 768-773, Dec 2015.
[4] P. Campr, M. Herbig, J. Vaněk and J. Psutka, "Sports video classification in continuous TV broadcasts," 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, pp. 648-652, 2014.
[5] V. Ellappan and R. Rajasekaran, "Event Recognition and Classification in Sports Video," 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), Tindivanam, pp. 182-187, 2017.
[6] F. Cricri et al., "Sport Type Classification of Mobile Videos," in IEEE Transactions on Multimedia, vol. 16, no. 4, pp. 917-932, June 2014.
[7] Alexander Filonenko, Laksono Kurnianggoro, Kang-Hyun Jo, "Smoke Detection on Video Sequences Using Convolutional and Recurrent Neural Networks", International Conference on Computational Collective Intelligence, pp. 558-566, 2017.
[8] Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y., "On the properties of neural machine translation: encoder-decoder approaches", computing research repository (2014).
[9] M. A. Russo, A. Filonenko and K.H. Jo, "Sports Classification in Sequential Frames Using CNN and RNN", ICT-Robot, Busan Korea, September 6-8, 2018.
[10] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li FeiFei. "Imagenet: A large-scale hierarchical image database." In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248-255. Ieee, 2009.
[11] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Springer, Cham, 2014.
[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556. 2014 Sep 4.